# Assignment 10.3 Step 2 of Final Project

Basak Atanu

05-19-2022

## Data Preparation for Exploratory Data Analysis

```
setwd("C:\\Users\\atanu\\Documents\\BellevueUniversity_MSDS\\DSC520\\Loan Defaulter Data")
default_fin <- read.csv("Default_Fin.csv")
head(default_fin)
```

```
##   Index Employed Bank.Balance Annual.Salary Defaulted.
## 1     1        1      8754.36     532339.56          0
## 2     2        0      9806.16     145273.56          0
## 3     3        1     12882.60     381205.68          0
## 4     4        1      6351.00     428453.88          0
## 5     5        1      9427.92     461562.00          0
## 6     6        0     11035.08      89898.72          0
```

**This data is related to defaulters, this gives individual's information like if the applicant is employed or not, their bank balance annual salary and if the application defaulted.**

```
setwd("C:\\Users\\atanu\\Documents\\BellevueUniversity_MSDS\\DSC520\\Loan Defaulter Data")
loan_data <- read.csv("loan_data.csv")
summary(loan_data)
```

```
##  credit.policy     purpose             int.rate       installment
##  Min.   :0.000   Length:9578        Min.   :0.0600   Min.   : 15.67
##  1st Qu.:1.000   Class :character   1st Qu.:0.1039   1st Qu.:163.77
##  Median :1.000   Mode  :character   Median :0.1221   Median :268.95
##  Mean   :0.805                      Mean   :0.1226   Mean   :319.09
##  3rd Qu.:1.000                      3rd Qu.:0.1407   3rd Qu.:432.76
##  Max.   :1.000                      Max.   :0.2164   Max.   :940.14
##  log.annual.inc        dti              fico       days.with.cr.line
##  Min.   : 7.548   Min.   : 0.000   Min.   :612.0   Min.   :  179
##  1st Qu.:10.558   1st Qu.: 7.213   1st Qu.:682.0   1st Qu.: 2820
##  Median :10.929   Median :12.665   Median :707.0   Median : 4140
##  Mean   :10.932   Mean   :12.607   Mean   :710.8   Mean   : 4561
##  3rd Qu.:11.291   3rd Qu.:17.950   3rd Qu.:737.0   3rd Qu.: 5730
##  Max.   :14.528   Max.   :29.960   Max.   :827.0   Max.   :17640
##    revol.bal         revol.util    inq.last.6mths    delinq.2yrs
##  Min.   :      0   Min.   :  0.0   Min.   : 0.000   Min.   : 0.0000
```

```
##  1st Qu.:   3187   1st Qu.: 22.6   1st Qu.: 0.000   1st Qu.: 0.0000
##  Median :   8596   Median : 46.3   Median : 1.000   Median : 0.0000
##  Mean   :  16914   Mean   : 46.8   Mean   : 1.577   Mean   : 0.1637
##  3rd Qu.:  18250   3rd Qu.: 70.9   3rd Qu.: 2.000   3rd Qu.: 0.0000
##  Max.   :1207359   Max.   :119.0   Max.   :33.000   Max.   :13.0000
##     pub.rec        not.fully.paid
##  Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000
##  Mean   :0.06212   Mean   :0.1601
##  3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :5.00000   Max.   :1.0000
```

**This dataset gives the loan details like the interest rate, fico of the customer, type of the loan, annual income along with fully paid or not flag.**

```
setwd("C:\\Users\\atanu\\Documents\\BellevueUniversity_MSDS\\DSC520\\Loan Defaulter Data")
application_data <- read.csv("application_data.csv")
```

**This data set is about loan application where Target field having 1 means the applicant have difficulty while paying for the loan and also have more than x day late payment.**

**Below are the list of Questions, that we are planning to answer using this data.**

**1. What attributes affect loan default and what are some major reasons behind it?**

**2. Is there any co-realation between different attributes of loan default data and gereral loan data?**

**3. I think, Income having a direct effect on loan default, because low income could cause default for loan payment. is it true?**

**4. Can I predict if the loan will go to default if I have employment, annual salary and bank balance information?**

**5. Does high fico socre give lower interest retes for loan?.**